



Getting the Performance Out Of High Performance Computing

Jack Dongarra
Innovative Computing Lab
University of Tennessee
and
Computer Science and Math Division
Oak Ridge National Lab
<http://www.cs.utk.edu/~dongarra/>

1



Getting the Performance Out Of ~~High Performance~~ Computing

Jack Dongarra
Innovative Computing Lab
University of Tennessee
and
Computer Science and Math Division
Oak Ridge National Lab
<http://www.cs.utk.edu/~dongarra/>

2


CNN.com/TECHNOLOGY

SEARCH

[Home Page](#)
[World](#)
[U.S.](#)
[Weather](#)
[Business & Finance](#)
[Sports & Hobbies](#)
[Politics](#)
[Law](#)
[Technology](#)
[Science & Space](#)
[Health](#)
[Entertainment](#)
[Travel](#)
[Education](#)
[Special Reports](#)



SERVICES
[Video](#)
[NewsWatch](#)
[E-Mail Services](#)
[CNN To Go](#)
SEARCH

Sherpa plans cyber cafe at Everest
 Monday, February 24, 2003 Posted: 2:40 PM EST (2040 GMT)

KATHMANDU, Nepal (Reuters) — The grandson of a Nepali sherpa in the first expedition to scale Mount Everest 50 years ago plans to set up the world's highest Internet cafe at the mountain's base camp.

Tensing Gyaltzen, whose grandfather, Gyaltzen Sherpa, was in the 1953 team that helped Sir Edmund Hillary and Tenzing Norgay reach the 8,850-meter (29,049-foot) summit, hopes to open the cafe next month to cash in on a flood of visitors for the anniversary.

Thousands of hikers and mountaineers pass through the base camp at 5,200 meters (17,400 feet) every year and many expeditions carry satellite phones into the Himalayas to run Web sites about their efforts and contact friends and family at home.

Otherwise, the nearest phones are a four-day trek away.

Gyaltzen, waiting for government permission to go ahead, will use radio and satellite links and solar and generator power.

Money from the cafe will go to a project to clear Everest of the hundreds of tons of garbage left behind every year.

Nepal has eight of the world's 14 highest mountains. The tens of thousands of foreign tourists they attract annually are a major source of income for what is one of the world's poorest nations.




Everest will soon have what's billed as the world's highest Internet cafe.

Story Tools
[SEND THIS](#) [EMAIL THIS](#)
[PRINT THIS](#) [MOST POPULAR](#)

RELATED
[Nepal Tourism Board](#)

3



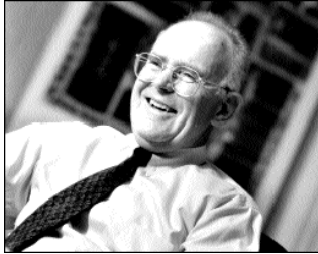
Getting the Performance into High Performance Computing

Jack Dongarra
 Innovative Computing Lab
 University of Tennessee
 and
 Computer Science and Math Division
 Oak Ridge National Lab
<http://www.cs.utk.edu/~dongarra/>

4

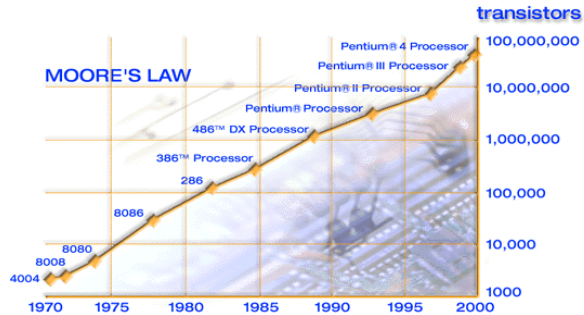


Technology Trends: Microprocessor Capacity



Gordon Moore (co-founder of Intel) predicted in 1965 that the transistor density of semiconductor chips would double roughly every 18 months.

2X transistors/Chip Every 1.5 years
Called “**Moore’s Law**”

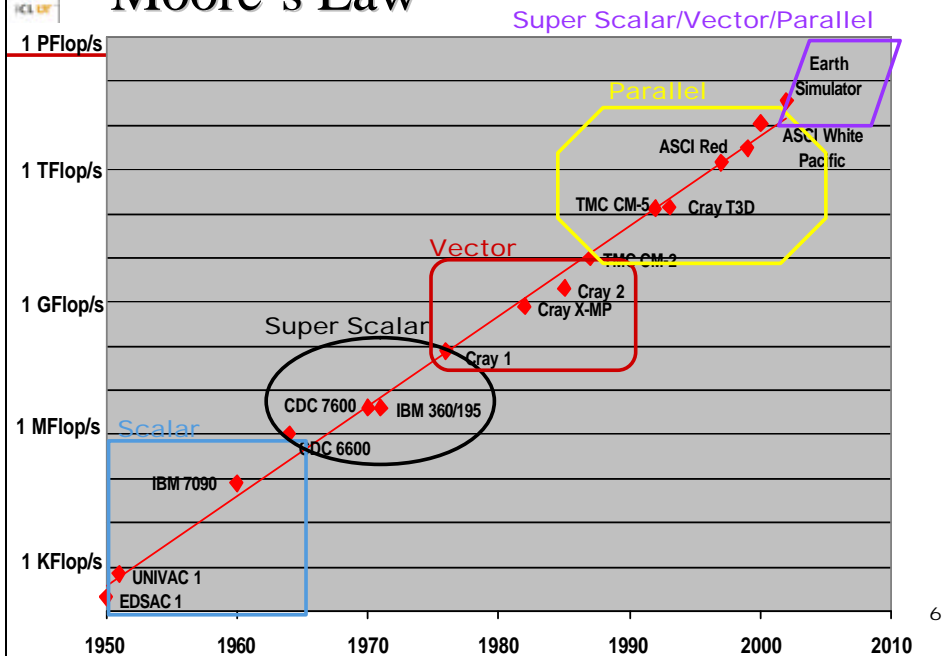


Microprocessors have become smaller, denser, and more powerful.
Not just processors, storage, internet bandwidth, etc

5



Moore’s Law



6

Next Generation: IBM Blue Gene/L and ASCI Purple

- ◆ **Announced 11/19/02**
 - **One of 2 machines for LLNL**
 - **360 TFlop/s**
 - **130,000 proc**
 - **Linux**
 - **FY 2005**

System (64 cabinets, 64x32x32)

Cabinet (32 Node boards, 8x8x16)

Node Board (32 chips, 4x4x2)
16 Compute Cards

Compute Card (2 chips, 2x1x1)

Chip (2 processors)

2.8/5.6 GF/s
4 MB

5.6/11.2 GF/s
0.5 GB DDR

90/180 GF/s
8 GB DDR

2.9/5.7 TF/s
256 GB DDR

180/360 TF/s
16 TB DDR

Plus
ASCI Purple
IBM Power 5 based
12K proc, 100 TFlop/s

To Be Provocative...
Citation in the Press, March 10th, 2008

National Report
The New York Times

DOE Supercomputers Sit Idle
WASHINGTON, Mar. 10, 2008

GAO reports that after almost 5 years of effort and several hundreds of M\$'s spent at the DOE labs, the high performance computers recently purchased did not meet users' expectation and are sitting idle...Alan Laub head of the DOE efforts

How could this happen?

- **Complexity of programming these machines were underestimated**
- **Users were unprepared for the lack of reliability of the hardware and software**
- **Little effort was spent to carry out medium and long term research activities to solve problems that were foreseen 5 years ago in the areas of applications, algorithm, middleware, programming models, and computer architectures, ...⁸**



Software Technology & Performance

- ◆ Tendency to focus on the hardware
- ◆ Software required to bridge an ever widening gap
- ◆ Gaps between potential and delivered performance is very steep
 - Performance only if the data and controls are setup just right
 - Otherwise, dramatic performance degradations, very unstable situation
 - Will become more unstable as systems change and become more complex
- ◆ Challenge for applications, libraries, and tools is formidable with Tflop/s level, even greater with Pflops, some might say insurmountable.

9



Linpack (100x100) Analysis, The Machine on My Desk 12 Years Ago and Today

- ◆ Compaq 386/SX20 SX with FPA - .16 Mflop/s
- ◆ Pentium IV - 2.8 GHz - 1317 Mflop/s
- ◆ 12 years → we see a factor of ~ 8231
 - Doubling in less than 12 months, for 12 years
- ◆ Moore's Law gives us a factor of 256.
- ◆ How do we get a factor > 8000?
 - Clock speed increase = 128x
 - External Bus Width & Caching -
 - 16 vs. 64 bits = 4x
 - Floating Point -
 - 4/8 bits multi vs. 64 bits (1 clock) = 8x
 - Compiler Technology = 2x
- ◆ However the potential for that Pentium 4 is 5.6 Gflop/s and here we are getting 1.32 Gflop/s
 - Still a factor of 4.25 off of peak

❖ Complex set of interaction between

- Application
- Algorithms
- Programming language
- Compiler
- Machine instructions
- Hardware

❖ Many layers of translation from the application to the hardware

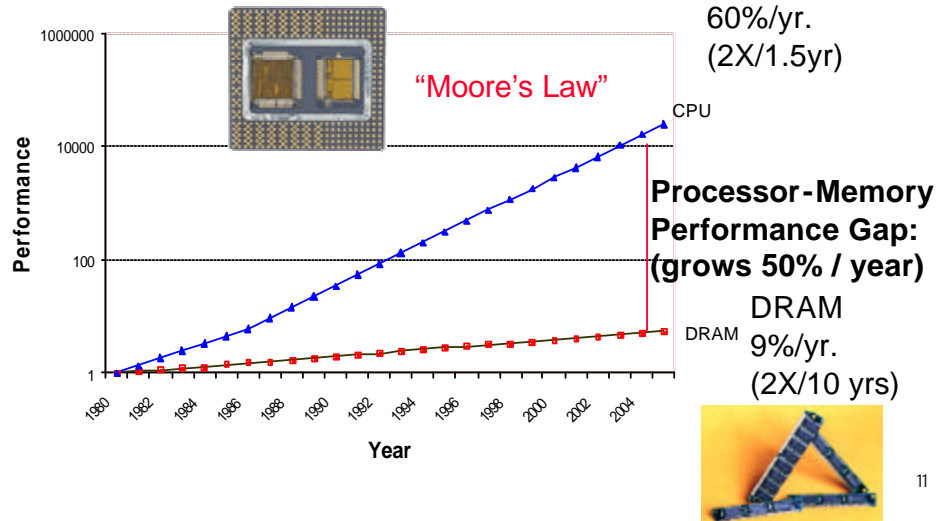
❖ Changing with each generation

10



Where Does Much of Our Lost Performance Go? or Why Should I Care About the Memory Hierarchy?

Processor-DRAM Memory Gap (latency)



Optimizing Computation and Memory Use

◆ Computational optimizations

- Theoretical peak: $(\# \text{ fpus}) * (\text{flops/cycle}) * \text{cycle time}$
 - Pentium 4: $(1 \text{ fpu}) * (2 \text{ flops/cycle}) * (2.8 \text{ Ghz}) = 5600 \text{ MFLOP/s}$

◆ Operations like:

- $y = a x + y$: 3 operands (24 Bytes) needed for 2 flops;
5600 Mflop/s requires 8400 MWord/s bandwidth from memory

◆ Memory optimization

- Theoretical peak: $(\text{bus width}) * (\text{bus speed})$
 - Pentium 4: $(32 \text{ bits}) * (533 \text{ Mhz}) = 2132 \text{ MB/s} = 266 \text{ MWord/s}$

Off by a factor of 30 from what's required to drive the processor from memory to peak performance

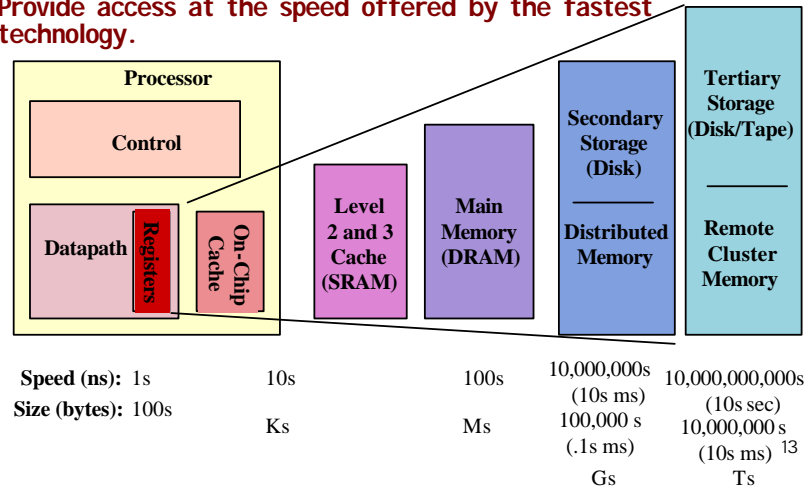
12



Memory Hierarchy

◆ **By taking advantage of the principle of locality:**


- Present the user with as much memory as is available in the cheapest technology.
- Provide access at the speed offered by the fastest technology.




Tool To Help Understand What's Going On In the Processor

- ◆ **Complex system with many filters**
- ◆ **Need to identify bottlenecks**
- ◆ **Prioritize optimization**
- ◆ **Focus on important aspects**

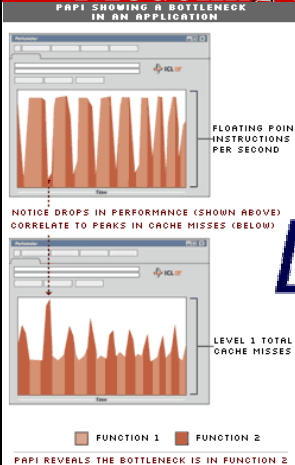
Tools for Performance Analysis, Modeling and Optimization



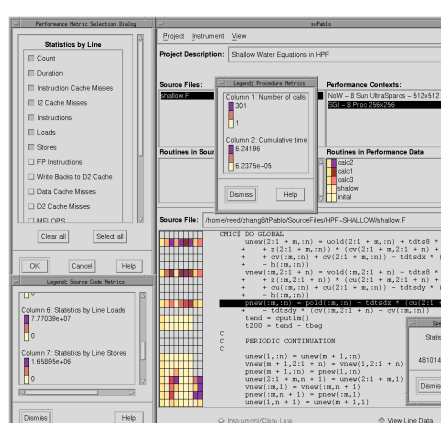
PAPI



Pablo
Scalable Performance Tools



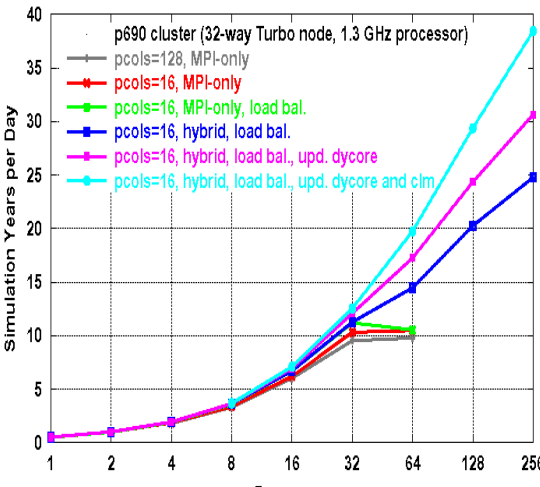
ROSE: Compiler Framework
Recognition of high-level abstractions
Specification of Transformations



Example PERC on Climate Model

- ♦ Interaction with the SciDAC Climate development effort
- ♦ Profiling
 - Identifying performance bottleneck and prioritizing enhancements
- ♦ Evaluation of code over time 9 month period
- ♦ Produced 400% improvement via decreased overhead and increased scalability

Performance Evolution of NCAR Community Atmospheric Model CAM2.0, EUL dynamical core, T42L26



Processors	pcols=128, MPI-only	pcols=16, MPI-only	pcols=16, MPI-only, load bal.	pcols=16, hybrid, load bal.	pcols=16, hybrid, load bal., upd. dycore	pcols=16, hybrid, load bal., upd. dycore and clm
1	0.5	0.5	0.5	0.5	0.5	0.5
2	1.0	1.0	1.0	1.0	1.0	1.0
4	2.0	2.0	2.0	2.0	2.0	2.0
8	4.0	4.0	4.0	4.0	4.0	4.0
16	8.0	8.0	8.0	8.0	8.0	8.0
32	10.0	10.0	10.0	10.0	10.0	10.0
64	10.0	10.0	10.0	10.0	10.0	10.0
128	10.0	10.0	10.0	10.0	10.0	10.0
256	10.0	10.0	10.0	10.0	10.0	10.0



Signatures: Key Factors in Applications and System that Affect Performance

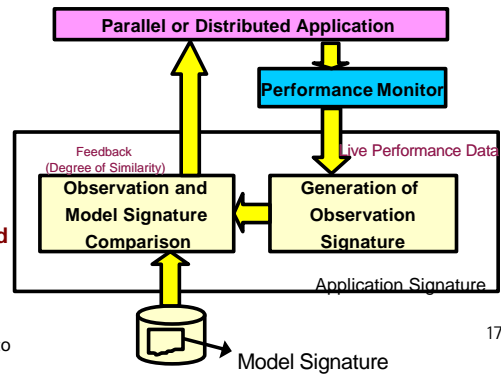
◆ Application Signatures

- Characterization of operations needed to be performed by application
- Description of application demands on resources
- **Algorithm Signatures**
 - Opts counts
 - Memory ref patterns
 - Data dependencies
 - I/O characteristics
- **Software Signatures**
 - Sync points
 - Thread level parallelism
 - Inst level parallelism
 - Ratio of mem ref to flpt ops
- Predict application behavior and performance

- Execution signature
- combine application and machine signatures to provide accurate performance models

◆ Hardware Signatures

- Performance capabilities of Machine
- Latencies and bandwidth of memory hierarchy
 - Local to node & to remote node
- Instruction issue rates
- Cache size
- TLB size



17



Algorithms vs Applications

	Lattice Gauge (QCD)	Quantum Chemistry	Weather Simulation	Comp Fluid Dynamics	Adjustment of Geodetic Networks	Inverse Problems	Structural Mechanics	Electronic Device Simulation	Circuit Simulation
Sparse Linear System Solvers	X			X	X		X	X	X
Linear Least Squares					X	X			
Nonlinear Algebraic System Solvers				X			X	X	X
Sparse Eigenvalue Problems	X	X					X		
FFT			X	X		X			
Rapid Elliptic Problem Solvers			X	X				X	
Multigrid Schemes				X				X	
Stiff ODE Solvers		X							X
Monte Carlo Schemes	X							X	
Integral Transformations		X							

18

From: Supercomputing Tradeoffs and the Cedar System, E. Davidson, D. Kuck, D. Lawrie, and A. Sameh, in High-Speed Computing, Scientific Applications and Algorithm Design, Ed R. Wilhelmson, U of I Press, 1986.

Update to Sameh's Table?

Application Performance Matrix

<http://www.krellinst.org/matrix/>



◆ Next step by looking at:

- Application Signatures
- Algorithms choices
- Software profile
- Architecture (Machine)

◆ Data mine to extract information

◆ Need signatures for A³S

19

Performance Tuning

◆ Motivation: performance of many applications dominated by a few kernels

◆ Conventional approach: handtuning by user or vendor

- Very time consuming and tedious work
- Even with intimate knowledge of architecture and compiler, performance hard to predict
- Growing list of kernels to tune
- Must be redone for every architecture, compiler
- Compiler technology often lags architecture
- Not just a compiler problem:
 - Best algorithm may depend on input, so some tuning at run-time.
 - Not all algorithms semantically or mathematically equivalent

20



Automatic Performance Tuning to Hide Complexity

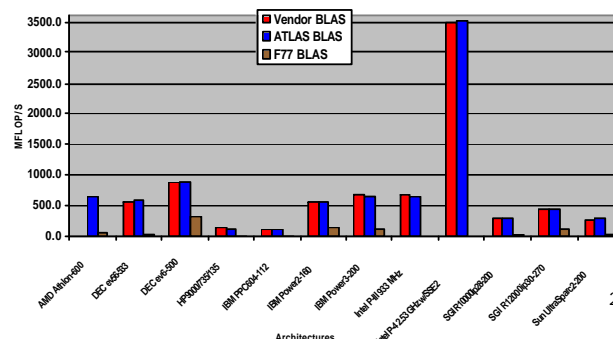
- ♦ **Approach: for each kernel**
 1. **Identify and generate a space of algorithms**
 2. **Search for the fastest one, by running them**
- ♦ **What is a space of algorithms?**
 - Depending on kernel and input, may vary
 - instruction mix and order
 - memory access patterns
 - data structures
 - mathematical formulation
- ♦ **When do we search?**
 - Once per kernel and architecture
 - At compile time
 - At run time
 - All of the above

21



Some Automatic Tuning Projects

- ♦ **ATLAS** (www.netlib.org/atlas) (Dongarra, Whaley)
used in Matlab and many SciDAC and ASCI projects
- ♦ **PHIPAC** (www.icsi.berkeley.edu/~bilmes/hipac) (Bilmes, Asanovic, Vuduc, Demmel)
- ♦ **Sparsity** (www.cs.berkeley.edu/~yelick/sparsity) (Yelick, Im)
- ♦ **Self Adapting Linear Algebra Software (SALAS)**
(Dongarra, Eijkhout, Gropp, Keyes)
- ♦ **FFTs and Signal Processing**
 - **FFTW** (www.fftw.org)
 - Won 1999 Wilkinson Prize for Numerical Software
 - **SPIRAL** (www.ece.cmu.edu/~spiral)
 - Extensions to other transforms, DSPs
 - **UHFFT**
 - Extensions to higher dimension, parallelism



22



Futures for High Performance Scientific Computing

- ♦ Numerical software will be adaptive, exploratory, and intelligent
- ♦ Determinism in numerical computing will be gone.
 - After all, its not reasonable to ask for exactness in numerical computations.
 - Reproducibility at a cost
- ♦ Importance of floating point arithmetic will be undiminished.
 - 16, 32, 64, 128 bits and beyond.
- ♦ Reproducibility, fault tolerance, and auditability
- ♦ Adaptivity is a key so applications can effectively use the resources.

23



Citation in the Press, March 10th, 2008

National Report

The New York Times

DOE Supercomputers

Live up to Expectation

WASHINGTON, Mar. 10, 2008

GAO reported today that after almost 5 years of effort and several hundreds of M\$'s spent at DOE labs, the high performance computers recently purchased have exceeded users' expectation and are helping to solve some of our most challenging problems.

Alan Laub head of DOE's HPC efforts reported today at the annual meeting of the SciDAC PI



How can this happen?

- ♦ Close interactions of with the applications and the CS and Math ISIC groups
- ♦ Dramatic improvements in adaptability of software to the execution environment
- ♦ Improved processor-memory bandwidth
- ♦ New large-scale system architectures and software
 - Aggressive fault management and reliability
- ♦ Exploration of some alternative architectures and languages
 - Application teams to help drive the design of new architectures

24



With Apologies to Gary Larson...



- ◆ SciDAC is helping
- ◆ Teams are developing the scientific computing software and hardware infrastructure needed to use terascale computers and beyond.

25